

Operations Research – Coursework

Instructions

This piece of coursework requires Python. As part of the coursework, you will be asked to submit your code and a short report as deliverables.

In the report, note the steps taken to solve each question, as well as include any discussion and observation if the question requires one.

The code should be annotated and commented to clearly show what question is being answered. Please, make sure your code executes fully before submission.

Both your code and accompanying report should have your CID clearly visible. Submit through blackboard by **5pm Thursday 5th December**. Use the following nomenclature:

- YOURCID-Report.pdf
- YOURCID-Code.zip

The zip file should include the Jupyter notebooks used to answer the questions and any supporting data or other files you should have used.

You have been provided a dataset of construction costs of homes alongside other home features, etc. You are tasked to prepare a regression model to predict apartment rent.

Exploratory Data Analysis

1. Load the dataset using pandas, and calculate the mean, standard deviation, and percentiles for each of the numerical features in the dataset. Calculate the number of apartments at each city and make a bar chart of the average apartment rent price in each city.

[5 marks]

2. Analyse the data in terms of apartment size and costs per city. Determine any outliers in the dataset. If so, filter the data as to remove the outliers from the dataset and repeat step 1.

Note: Scatter or box-and-whisker plots may prove useful.

[5 marks]

Use apartment size as the only feature for the following questions:

Regression

3. Apply the least-square method to calculate the intercept and slope for the linear regression equation that estimates apartment rent. Discuss your findings and accuracy of the regression model.

[15 marks]

4. Derive the gradient descent update rule and perform the first two iterations of it using $\alpha_0 = -10000$ and $\beta_0 = 20$, and learning rate $\eta = 10^{-9}$ as the initial parameters.

[15 marks]

Learning

Using the code provided to you in the Gradient Descent Linear Regression notebook:

5. Implement a gradient descent model to optimise the intercept and slope for 30 iterations. Use the same initial conditions as in **part 4**. **Plot alongside the calculated values of part 3**. What are the final slope and intercept?

[15 marks]

6. Implement your code in the previous question to analyse how the best learning rate affects the effectiveness of the gradient descent algorithm. Discuss the effect of the learning rate on the algorithm's accuracy.

[10 marks]

7. Discuss the application of gradient descent and suggest alternative termination conditions, instead of number of iterations. Modify the Gradient Descent code to apply an alternative termination condition.

[15 marks]

Application

8. Use sk-learn to train a linear regression model and test its accuracy. Determine the accuracy of your model in predicting the rent costs by city.

[10 marks]

9. What insights do you gather based on the outputs of **part 8**? Discuss the performance of your model, indicating limitations and providing recommendations, if any, to improve model performance.

[25 marks]